5-28-2020

Mamma Mia, Here You Blow Again! My My, How Can We Protect You? – The Development of a Threat Model for Whistleblower Anonymity

Julia DeRango

Ethan Lederer

Meghan Lennon

Jacob Young
jayoung@fsmail.bradley.edu

Mamma Mia, Here You Blow Again! My My, How Can We Protect You? – The Development of a Threat Model for Whistleblower Anonymity

Julia T. DeRango
Bradley University
jderango@mail.bradley.edu

Ethan L. Lederer
Bradley University
elederer@mail.bradley.edu

Meghan K. Lennon
Bradley University
mlennon@mail.bradley.edu

Jacob A. Young
Bradley University
jayoung@fsmail.bradley.edu

ABSTRACT

This research-in-progress paper focuses on the development of a threat model for maintaining whistleblower anonymity. Citizens have the obligation to uphold ethical behavior and report behavior that violates their personal and professional code of ethics. Whistleblowers are those who release information to expose the wrongdoings of a company or individual. When choosing how to report, the whistleblower must consider various reporting channels. However, due to the likelihood of retaliation, whistleblowers must also be aware of threats to their anonymity. Therefore, anonymous whistleblowers must take actions to protect themselves from threat actors and threat vectors that might expose their identity. With these issues in mind, our goal is to develop a comprehensive threat model for anonymous whistleblowing, as well as provide whistleblowers with a systematic approach to maintaining their anonymity throughout the whistleblowing process.

Keywords

Whistleblowing, threat modeling, anonymity

INTRODUCTION

To effectively curb unethical or illegal behavior throughout society, it is imperative that a reporting method exists where the whistleblower can achieve total anonymity. In this research-in-progress paper, we outline our efforts to develop a guide to assist someone in blowing the whistle anonymously. First, we discuss key terms in the context of whistleblowing, such as identity, anonymity, confidentiality, and credibility. Second, we discuss various reporting channels available to whistleblowers. Third, we provide an overview of threat modeling methodologies and identify likely threat actors and threat vectors. We rely on Shostack (2014) as we develop a comprehensive threat model that clearly illustrates all the possible threat actors and threat vectors. The use of STRIDE, Privacy Impact Assessments, and LINDDUN will assist in this process. Lastly, we discuss the planned outputs of our research efforts. Our goal is to protect whistleblowers from retaliation by ensuring they remain truly anonymous when reporting wrongdoing.

KEY TERMS

Identity

As technology continues to evolve, so does the list of the many ways someone can be identified. According to Marx (1999), the seven types of identity knowledge that can reveal one's identity are: legal name, locatability, pseudonyms linked to name or location, pseudonyms not linked to name or location, pattern knowledge, social categorization, and symbols of eligibility/non-eligibility.

Anonymity

Given Marx's typology, "to be fully anonymous means that a person cannot be identified according to any of the seven dimensions of identity knowledge" (Marx, 1999, p. 100). For our purposes, we have adopted the definition provided by Pfitzmann & Köhntopp (2001), which states that "anonymity is the state of being not identifiable within a set of subjects, the anonymity set." In the context of whistleblowing, the anonymity set consists of all the individuals who could have detected

the reported wrongdoing. The larger the set, the stronger an individual's anonymity becomes, as there are more potential sources for the report. It is important to note that the goal of staying anonymous is not to prevent others observing an act committed by a whistleblower, but rather to disassociate their identity from the act (Anonymous, 1998).

Confidentiality

Confidentiality is often erroneously used in place of anonymity, and the distinction between them is crucial for whistleblowers to understand. Confidentiality is different than anonymity "in that the source of a comment is known to a few, but the identity of the source is not further revealed. Thus, the source is absent to all with anonymity; however, confidentiality is a condition in which the source can be connected to his or her comments by some who agree not to reveal the source to others" (Anonymous, 1998, p. 383). Therefore, confidentiality should not be considered adequate to protect whistleblowers since at least one other person is aware of the whistleblowers' identity.

Credibility

According to Metzger (2007, p. 2078), credibility refers to "the believability of some information and/or its source." It is widely recognized across society that anonymous sources are generally perceived as less credible (Rains & Scott, 2007). Signaling theory suggests that claims can be supported through one of two signaling methods (Donath, 2002). Conventional signals are traits or values that the sender does not have to possess, such as someone claiming to have incriminating inside information against an organization. Alternatively, assessment signals are directly related to the characteristics of the sender and are perceived to be more credible than conventional methods. Therefore, when whistleblowers conceal their persuasive assessment signals, such as their identity, they might be perceived as less credible.

WHISTLEBLOWING REPORTING CHANNELS

The likelihood of retaliation against whistleblowers is high (Ethics & Compliance Initiative, 2019). Therefore, once a whistleblower has decided to report their concerns, they must be conscious of how certain reporting channels might jeopardize their anonymity and lead to retaliation. When evaluating various reporting channels, the whistleblower must consider their organization's policies, the severity of the situation, and above all, their anonymity. Reporting channels can be classified as either internal or external. All channels have significant risks that must be addressed.

Internal

Internal reporting channels may be convenient; however, the whistleblower puts their anonymity at risk when complying with certain company policies. Given the threat of retaliation, we argue that whistleblowers must view maintaining their anonymity as the most important aspect of the reporting process. Unfortunately, many internal channels do not fully support anonymity, which only encourages whistleblowers to seek out external channels. For example, the internal reporting channel that may produce the most risk is an open-door policy. An open-door policy is a commonly accepted workplace practice that gives an internal reporter a "safe space" to report unethical behavior. Although this may afford the whistleblower credibility, the idea of an open-door policy completely defeats the idea of anonymity. Not only does this give the company complete control over the whistleblower, but it may also discourage internal reporting altogether. If the company does offer an internal reporting channel, such as a "secure" website, the whistleblower must take steps to certify its security and anonymity. If the reporting channel requires any sort of personal information such as email, company id, or name, the whistleblower should instead seek an alternate channel.

External

Anonymity is a large factor in making a report secure and successful. If the whistleblower is unable to maintain their anonymity, there is no way to acquire it again, which puts them at risk of retaliation. Although there are many risks associated with internal reporting, the whistleblower must accept that there is an abundance of risk with external reporting as well. If the whistleblower chooses an external reporting channel, such as directly contacting the media or a government agency, they are at risk of exposure. The reaction to news media is often explosive and once published, the whistleblower will no longer have control of their story. If reporting to inexperienced members of the media, the whistleblower's identity might be exposed through the publishing of documents or incriminating details.

THREAT MODELING

The purpose behind threat modeling is to systematically assess a system or process to ensure nothing is overlooked. There are several methodologies available for threat modeling, such as STRIDE and attack libraries. There are also other privacy tools,

such as Solove's Taxonomy of Privacy, Privacy Impact Assessments, and LINDDUN. In this section, we briefly discuss each methodology, then we identify several threat actors and threat vectors relevant to the whistleblowing context.

Methodologies

STRIDE is a mnemonic that focuses on six primary threats: spoofing, tampering, repudiation, information disclosure, denial of service, and elevation of privilege. Attack libraries are used to find potential threats against a system that is being constructed. Depending on the situation, attack libraries are a more detailed list to prepare for an attack and are preferred over STRIDE.

Daniel Solove proposed the Taxonomy of Privacy (Solove, 2006). The first element is *information collection* through forms of surveillance and interrogation. The second element is *information processing* through the aggregation of data, the identification of a person involved, or simple insecurity of disclosing information. The third element is *information dissemination* where confidentiality is breached, and personal matters are disclosed, exposed, and easily accessible. The last element is *invasion*, either through unwanted intrusion or decisional interference.

Privacy Impact Assessments (PIA) are used to evaluate how a stakeholder's privacy might be impacted by various initiatives or projects. The process of conducting a PIA is to describe the project and data flows, then analyze against information privacy principles, other privacy dimensions, and the privacy control environment. Once the assessment has been completed, appropriate privacy and security recommendations can be proposed and adopted.

LINDDUN is a mirror of the STRIDE mnemonic device that is used to find threats to privacy (Deng, Wuyts, Scandariato, Preneel, & Joosen, 2011). It is comprised of the following properties of privacy violations: linkability, identifiability, non-repudiation, detectability, disclosure of information, content unawareness, and policy and consent noncompliance. LINDDUN is a complete process that can be used in place of, or along with, STRIDE.

Threat Actors

Threat actors are individuals or entities who might jeopardize the safety and security of another. Accounting for all parties who could be included in the whistleblowing process can negatively impact anonymity. In the whistleblowing context, potential threat actors could include the affected parties, news media, competitors, and government agencies.

The accused, if innocent, will want to clear their name and distance themselves from the issue. If the accused is guilty, they might want to hide and dissociate themselves from the issue. If the allegations are made public, innocent and accused parties might be motivated to retaliate and seek to identify the whistleblower. Proactive organizations will try to resolve the matter at hand before it creates a bigger problem, such as through the media. However, others might attempt to silence the whistleblower.

Even if a story is published in the news media by a reputable reporter, that does not prevent other news outlets from attempting to identify the whistleblower. Competitors are also a major threat actor because they are motivated to obtain an advantage for themselves. This could be achieved by reporting false allegations about a company or stealing legitimate reports to harm to another organization. Government agencies might also seek to uncover a whistleblower's identity, either to penalize a wrongdoer, or prevent certain information from coming to light.

We plan to adopt and develop personas for each threat actor according to the approach of Aucsmith, Dixon, & Martin-Emerson (2003), as reproduced by Shostack (2014). Each persona will be evaluated in terms of motivation and skill level.

Threat Vectors

A *threat vector* refers to a path or tool used by a threat actor. In the context of whistleblowing, these vectors could be used to pinpoint the whistleblower through various modes of identification.

Internal

Internal threat vectors can refer to a whistleblower jeopardizing their identity through their actions taken within the organization, such as accessing documents or sharing their concerns with a coworker. Internal whistleblowing activity can be detected through digital signatures, security logs, and time stamps.

For example, a *canary token* is a defense mechanism that can alert the owners of a digital resource if it is accessed by others. After Edward Snowden's disclosures in 2013, the U.S. Central Intelligence Agency (CIA) developed a new system called Scribbles. Its purpose was to catch future whistleblowers through a digital watermark that informs the CIA when a document

is opened. Further, a whistleblower's location or IP address can easily give away their identity, such as through social media, email, or even with a smartphone on standby.

Coworkers can also jeopardize a whistleblower's anonymity by exposing their colleague through communication, or even observation, and reporting them to their supervisors. Supervisors can also play a part in identifying their subordinates. For example, emails sent to the founder of Theranos by Tyler Shultz were used to identify him as a whistleblower.

External

External threat vectors are those not formally connected to the whistleblower's organization. This might include a family member, a trusted third party who decides to take concerns into their own hands. Friends are another possibility that the threat actor may confide in, revealing information that the companion may see as valuable to release to the public. Similarly, whistleblowers are unlikely to be adequately prepared to anonymously disclose concerns on social media.

Media outlets are highly motivated to break stories and reputable journalists have gone to great lengths to defend their sources. However, sharing documents with a member of the media can potentially jeopardize a whistleblower's anonymity. For example, Reality Winner was identified and arrested after unredacted copies of leaked documents marked with printer microdots were published by *The Intercept*. Microdots are semi-invisible dots placed on documents when printed so that its origin could later be traced if the document were to be shared.

FUTURE WORK

Our preliminary threat matrix is provided in Table 1. In addition to further development of a comprehensive threat model, we plan to extend our research into a practical guide for whistleblowers to follow when considering their options. Each whistleblower must first be informed of the various potential threats before they can evaluate their unique situation and determine the safest method of reporting their concern. Following a systematic approach will simplify a complicated, yet critical process, which should minimize mistakes that undermine whistleblower anonymity.

	Internal	**External**
Technical	• Activity logs [network, access control, etc.] • User information [usernames, passwords, etc.] • Reporting systems [phone hotline, web form, etc.]	• Networked communication [traffic analysis] • User information [usernames, passwords, etc.] • Reporting systems [phone hotline, web form, etc.]
Personal	• Communication [supervisor, coworkers, etc.] • Observable evidence gathering or reporting activities	• Communication [family, friends, attorneys, etc.] • Observable evidence gathering or reporting activities
Contextual	• Who was involved in the alleged wrongdoing [wrongdoer(s)/witness(es)]? • What did the alleged wrongdoing involve? • When did the alleged wrongdoing occur? • Where did the alleged wrongdoing take place?	• Who has the whistleblower contacted? • What actions has the whistleblower taken? • Where has the whistleblower visited [physical/digital]?

Table 1. Preliminary Threat Matrix

CONCLUSION

For society to effectively detect and reduce unethical behavior, whistleblowers must be provided a secure and anonymous channel to report their concerns. Our research aims to provide a cohesive and effective strategy for whistleblowers to report wrongdoing without compromising their identity. We intend to develop a threat model that considers all threat actors and threat vectors that might undermine whistleblower identity. With proper anonymity, secure reporting channels, and in-depth threat models, it is our hope that whistleblowers may start to feel more confident in their decision to expose wrongdoing.

ACKNOWLEDGMENTS

This research was supported by Bradley University's Center for Cybersecurity and the National Whistleblower Center.

REFERENCES

1. Anonymous. (1998). To Reveal or Not to Reveal: A Theoretical Model of Anonymous Communication. *Communication Theory*, *8*(4), 381–407.
2. Aucsmith, D., Dixon, B., & Martin-Emerson, R. (2003). *Threat Personas* (No. version 0.9).
3. Deng, M., Wuyts, K., Scandariato, R., Preneel, B., & Joosen, W. (2011). A privacy threat analysis framework: Supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, *16*(1), 3–32. https://doi.org/10.1007/s00766-010-0115-7
4. Donath, J. (2002). Identity and Deception in the Virtual Community. In P. Kollock & M. Smith (Eds.), *Communities in Cyberspace* (pp. 37–68). London, UK, UK: Routledge.
5. Ethics & Compliance Initiative. (2019). *Workplace Misconduct and Reporting: A Global Look*. Vienna, Virginia.
6. Marx, G. T. (1999). What's in a Name? Some Reflections on the Sociology of Anonymity. *The Information Society*, *15*(2), 99–112.
7. Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, *58*(13), 2078–2091. https://doi.org/10.1002/asi.20672
8. Pfitzmann, A., & Köhntopp, M. (2001). Anonymity, Unobservability, and Pseudonymity — A Proposal for Terminology. In H. Federrath (Ed.), *Designing Privacy Enhancing Technologies* (pp. 1–9). https://doi.org/10.1007/3-540-44702-4_1
9. Rains, S. A., & Scott, C. R. (2007). To Identify or Not to Identify: A Theoretical Model of Receiver Responses to Anonymous Communication. *Communication Theory*, *17*(1), 61–91. https://doi.org/10.1111/j.1468-2885.2007.00288.x
10. Shostack, A. (2014). *Threat Modeling: Designing for Security* (1st ed.; C. Long, ed.). Indianapolis: Wiley.
11. Solove, D. J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review*, *154*(3), 477–560.